Expert Systems with Applications 39 (2012) 7562–7573

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Plant species identification using digital morphometrics: A review

James S. Cope [a], David Corney [b,*], Jonathan Y. Clark [b], Paolo Remagnino [a], Paul Wilkin [c]

[a] Digital Imaging Research Centre, Kingston University, London, UK
[b] Department of Computing, University of Surrey, Guildford, Surrey, UK
[c] Royal Botanic Gardens, Kew, Richmond, Surrey, UK

ARTICLE INFO

Keywords:
Morphometrics
Shape analysis
Image processing
Plant science
Leaf
Flower
Taxonomy

ABSTRACT

Plants are of fundamental importance to life on Earth. The shapes of leaves, petals and whole plants are of great significance to plant science, as they can help to distinguish between different species, to measure plant health, and even to model climate change. The growing interest in biodiversity and the increasing availability of digital images combine to make this topic timely. The global shortage of expert taxonomists further increases the demand for software tools that can recognize and characterize plants from images. A robust automated species identification system would allow people with only limited botanical training and expertise to carry out valuable field work.

We review the main computational, morphometric and image processing methods that have been used in recent years to analyze images of plants, introducing readers to relevant botanical concepts along the way. We discuss the measurement of leaf outlines, flower shape, vein structures and leaf textures, and describe a wide range of analytical methods in use. We also discuss a number of systems that apply this research, including prototypes of hand-held digital field guides and various robotic systems used in agriculture. We conclude with a discussion of ongoing work and outstanding problems in the area.

1. Introduction

Plants form a fundamental part of life on Earth, providing us with breathable oxygen, food, fuel, medicine and more besides. Plants also help to regulate the climate, provide habitats and food for insects and other animals and provide a natural way to regulate flooding. A good understanding of plants is necessary to improve agricultural productivity and sustainability, to discover new pharmaceuticals, to plan for and mitigate the worst effects of climate change, and to come to a better understanding of life as a whole.

With a growing human population and a changing climate, there is an increasing threat to many ecosystems. It is therefore becoming increasingly important to identify new or rare species and to measure their geographical extent as part of wider biodiversity projects. Estimates of numbers of species of flowering plants (or angiosperms) vary from about 220,000 (Scotland & Wortley, 2003) to 420,000 (Govaerts, 2001).

The traditional approach to identifying species and their relationships is to train taxonomists who can examine specimens and assign taxonomic labels to them. However, there is a shortage of such skilled subject matter experts (a problem known as the "taxonomic impediment" e.g. Carvalho et al., 2007), as well as a limit on financial resources. Furthermore, an expert on one species or family may be unfamiliar with another. This has lead to an increasing interest in automating the process of species identification and related tasks. The development and ubiquity of relevant technologies, such as digital cameras and portable computers has bought these ideas closer to reality; it has been claimed that now is the "time to automate identification" (MacLeod, Benfield, & Culverhouse, 2010), and not just of plants. Arguing that we need to train more expert taxonomists, while also embracing new technologies, Quentin Wheeler writes that "[d]igital images are to morphological knowledge what the Gutenberg Press was to the written word" (Wheeler, 2004).

Botanists collect specimens of plants and preserve them in archives in herbaria. For example, the herbarium at the Royal Botanic Gardens, Kew in London houses over 7 million dried specimens,[1] some of which are more than 200 years old. These are annotated and sorted using the expert knowledge of the botanists, subject to revision over time. Herbarium collections can therefore be seen as major, structured repositories of expert knowledge. In order to improve access, these collections are increasingly being digitized to form databases with images that are annotated with species' names, collectors' names, dates, locations and so on. Other significant sources of knowledge include flora, taxonomic keys and monographs (see Table 1 for definitions).

* Corresponding author.
E-mail addresses: J.Cope@kingston.ac.uk (J.S. Cope), D.Corney@surrey.ac.uk (D. Corney), J.Y.Clark@surrey.ac.uk (J.Y. Clark), P.Remagnino@kingston.ac.uk (P. Remagnino), P.Wilkin@kew.org (P. Wilkin).

[1] http://apps.kew.org/herbcat/navigator.do.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.073

Table 1
Some botanical terminology. Note that some terms have different meanings in plant science compared to computer science or statistics. See also Fig. 2 for terms relating to the anatomy of a leaf.

Identification	Recognition of the identity of an organism. (Synonymous with *classification* in computer science and statistics.)
Classification	Grouping items based on similarity. (Synonymous with *cluster analysis* or *segmentation* in computer science and statistics.)
Nomenclature	Assigning names to organisms
Taxonomy	Identification, formal description and naming of organisms
Taxon	Group of organisms assumed to be a unit; e.g. a species
Taxonomic rank	Relative position in taxonomic hierarchy; e.g. "species", "family"
Dichotomous key	A binary tree that allows a user to identify members of a taxon through a series of questions
Flora	A book describing plant life in a particular geographic region
Monograph	A book providing a (near) complete description of a particular taxon, typically a genus
Taxonomic key	Structured series of questions used to identify specimens
Herbarium	A reference collection of preserved plant specimens.
Systematics	The taxonomic study of evolutionary origins and environmental adaptations
Cladistics	The study of the pathways of evolution, with the aim of identifying ancestor–descendant relationships
Phenetics	The study of relationships between organisms defined by the degree of physical similarity between them
Homology	The similarity of a structure in different organisms resulting from shared ancestry

In many cases, species (or higher taxa such as genera or families) can be distinguished by characters derived from their leaf or flower shape, or their branching structure. Shape is of course important in many other disciplines, and morphometric techniques are applied in structurally-based research in zoology, geology, archaeology and medicine, although these are beyond the scope of this review.

Morphometrics, the study of shape, has been applied to plants and their organs for many years. Leaves are readily apparent structures on many plants, and they are available for examination for much of the year in deciduous or annual plants or year-round in evergreen perennials, unlike more transient reproductive organs. As such, leaf characters, including those involving shape, have been used extensively in traditional text-based taxonomic keys for plant identification since the beginnings of botany. Examples of such studies include those on *Tilia* (Schneider, 1912), *Ulmus* (Melville, 1937, 1939) and *Betula* (Natho, 1959), but there are many more. To use such a key, which has been compiled by an expert in the group in question, the user makes a series of choices between contrasting statements, finally reaching a species name. Even given such a key, the user must make a number of judgements that require specific botanical knowledge, so these cannot be used naively. Further details on taxonomic keys can be found in Stace (Stace, 1992).

In recent years, high quality digital cameras have become ubiquitous, increasing interest in creating hand-held field guides. These are prototypes built around smart phones or personal digital assistants (PDAs) that are designed to allow a user in the field to photograph a specimen of interest and instantly receive information about it, such as the likely species name (see also Section 3). One advantage of such systems is that they require little infrastructure at the point of use, so can be used even in the least developed and most remote parts of the world. However, the scope of such systems is currently very limited, restricting their practical use.

A second consequence of cheap digital cameras and scanners is the creation of vast databases of plant images. For example, the Royal Botanic Gardens, Kew provides a digital catalogue of over 200,000 high-resolution images, with more being added continuously as part of an ongoing digitization project. We maintain an annotated list of botanical image sets online,[2] describing various publicly available sets, including images of single-leaves, herbarium specimens, and whole plants.

1.1. Challenges in botanical morphometrics

Although morphometrics and image processing are well-established and broad disciplines, botanical morphometrics presents some specific challenges. Here, we discuss some of these, including specimen deformations, unclear class boundaries, feature selection and terminology.

Leaves and flowers are non-rigid objects, leading to a variety of deformations. Many leaves have a three-dimensional nature, which increases the difficulty of producing good quality leaf images and also results in the loss of useful structure information. Archived specimens may also be damaged as they are dried and pressed, but even live specimens may have insect, disease or mechanical damage. Automated systems must be robust to such deformations, making soft computing and robust statistics highly attractive.

One source of confusion when botanists and computer scientists collaborate concerns terms such as "classify" and "cluster". In taxonomy, "classification" may be defined as the process of grouping individuals based on similarity, in order to define taxa such as species or genera (Stuessy, 2006). "Identification" is then the process of deciding to which of a number of pre-defined taxa a particular individual belongs. In computer science by contrast, "classification" refers to the assigning of an individual example to one of a finite number of discrete categories, whereas "clustering" refers to the discovery of groups within a set of individuals, based on similarities (Bishop, 2007, p.3). Care must be taken when using such terms to avoid confusion.

Any system that is concerned with distinguishing between different groups of plants must be aware of the large intra-class, and small inter-class variation that is typical of botanical samples (see Fig. 1). A number of classifiers have been developed that identify the species of a specimen from a digital image, as we discuss throughout this paper, and these must be robust to this challenge. Similar issues apply to the tasks of discovering how many groups exist in a set of examples, and what the class boundaries are. See Figs. 3 and 7 for further examples of the variety of leaf shapes found.

Distinguishing between a large number of groups is inherently more complex than distinguishing between just a few, and typically requires far more data to achieve satisfactory performance. Even if a study is restricted to a single genus, it may contain many species, each of which will encompass variation between its constituent populations. The flowering plant genus *Dioscorea*, for example, contains over 600 species (Govaerts, Wilkin, Raz, & Téllez-Valdés, 2010), so even single-genus studies can be very challenging. On a related note, the shape of leaves may vary continuously or discretely along a single stem as the leaves develop (known as leaf heteroblasty), which can further confound shape analysis unless careful attention is paid to sources of the specimens.

Different features are often needed to distinguish different categories of plant. For example, whilst leaf shape may be sufficient to

[2] http://www.computing.surrey.ac.uk/morphidas/ImageSets.html.